# Robotics Research Technical Report

eneratorium omnis laboris ex machina

PRELIMINARY IMPLEMENTATION OF A RATIO
IMAGE DEPTH SENSOR

by

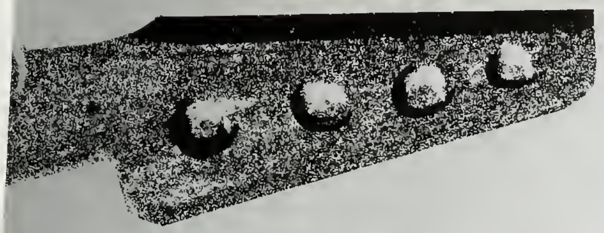
C. M. Bastuscheck

and

J. T. Schwartz

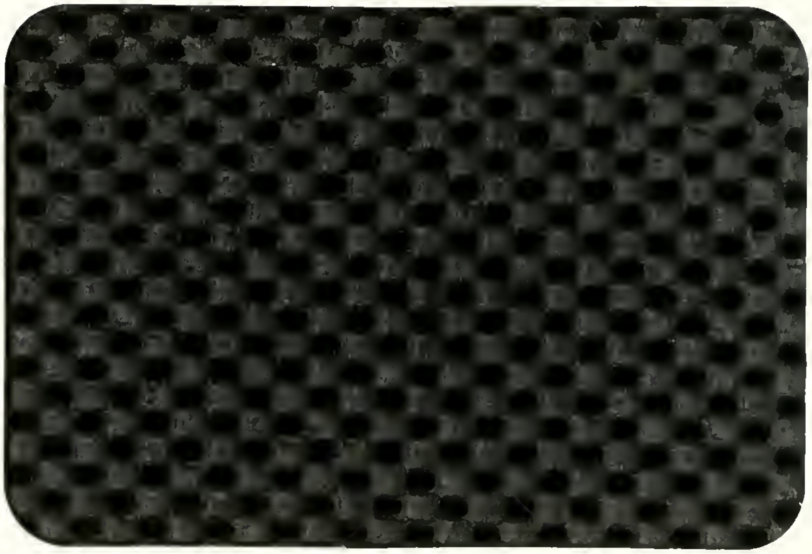
Technical Report No. 124
Robotics Report No. 28

June, 1984



New York University
Courant Institute of Mathematical Sciences

Computer Science Division
251 Mercer Street New York, N.Y. 10012



# PRELIMINARY IMPLEMENTATION OF A RATIO IMAGE DEPTH SENSOR

by

C. M. Bastuscheck

and

J. T. Schwartz

Technical Report No. 124
Robotics Report No. 28

June, 1984

New York University
Courant Institute of
Mathematical Sciences
251 Mercer Street
New York, NY 10012

# Preliminary Implementation of a Ratio Image Depth Sensor*

C. M. Bastuscheck and J. T. Schwartz

Robotics Activity, Dept. of Computer Science
Courant Institute of Mathematical Sciences, New York University

Abstract: We describe a novel variant of depth measurement by optical triangulation in which information is recorded simultaneously from an entire scene rather than point-by-point or plane-by-plane. The implementation uses standard components to form a 7 bit depth image in approximately 100 seconds.

## I. Introduction

The acquisition of geometric data from 3D scenes is an important issue for computer vision. Considerable effort has gone into the development of various methods of extracting geometric information from 2D images of scenes as well as into the development of various range finding techniques to record depth information directly; see [1,2] for recent reviews. The most successful range finders of interest to robotics at present are a technique of dynamic stereo [3,4] and several plane-of-light triangulation schemes [5,6,7] which are able to record arbitrary shapes with high resolution. However, even these are too slow to be immediately useful in robotics applications. The purpose of this paper is to introduce a variation of optical triangulation in which geometric information is gathered from an entire scene at once rather than plane-by-plane or point-by-point. Properly engineered, this new method [9] promises to speed up the acquisition of range information considerably.

The remainder of this introductory section reviews some of the many approaches to the problem of acquiring geometric information from a scene. The second section describes our novel structured light method. The third section describes an elementary implementation of this Ratio Image method which we have used to test the behavior of our theoretical assumptions in practice, and shows a 'depth image' made using this implementation. It also presents the results of some preliminary measurements designed to elucidate the sources of error in the measurements. The fourth and final section discusses factors which limit the resolution of images made with this method.

The amount of spatial information which can be extracted from an image such as is made by a camera is distinctly limited. While it is possible to exploit occlusion cues [10] or texture [11] to obtain limited spatial relations between

---

* Work on this paper has been supported in part by the Office of Naval Research grant N00014-82-K-0381, and by grants from the Digital Equipment Corporation, the System Develop-

objects or features in the original scene, it is impossible to establish the corresponding absolute geometric positions using only one image of a 3D scene. An exception to this may be the technique of 'shape from shading' [12] which can allow a good guess about local geometry within a scene, but even here it is difficult to reconstruct all 3D information from a single 2D image.

By using more than one image of a scene it is in principle possible to determine the geometrical relations in the original 3D scene for those regions appearing in more than one projection. Much effort has been devoted to computer stereo vision (see e.g. [13]) and to studies of optical flow (e.g. [14]). In both approaches the geometry of a 3D scene is deduced by correlating the locations of corresponding points in images taken from different known locations. Two difficulties must be overcome to do this. First, one must identify corresponding points in images having very low resolution compared to human vision, and secondly one must face an inherent trade-off between large camera separation (which increases the geometrical resolution) and small separation (which makes it easier to identify corresponding points). The technique of photometric stereo [15,16] avoids these difficulties by using several images taken from the same viewpoint but under different (known) lighting conditions; however, all these techniques require substantial amounts of computation.

A more direct approach to finding depth by the use of contrived lighting is the method of optical triangulation, developed by Will and Pennington and by Shirai [17] over 15 years ago. In this procedure a computer with a television camera records the location of points illuminated by a vertical plane of light projected obliquely across the field of view of the camera. The location of any illuminated point (bright pixel in the image) is determined by the intersection of the known plane of light and the ray from the camera corresponding to the illuminated pixel. Information from the entire scene is acquired by moving the plane of light through a number of different angles and recording the locations of illuminated points for each angle.

One difficulty with triangulation methods is that information is available only for those regions of the 3D scene which are both illuminated and visible to the camera. Thus some information (such as the depth of a narrow hole facing the camera) can never be known, even when two or more projectors are used. This deficiency has been overcome by the use of laser range finders which scan a laser spot over the 3D scene and detect the light reflected back over the same optical path as the incident ray. There are two such methods, one a modulation technique [18] in which the range is determined by the difference in modulation phase between the light source and the light returning from the work scene, and the other that of pulse time of flight [19]. Both methods encounter difficulty with the large dynamic range of the reflected light, with secondary reflections within the work scene, and with low signal-to-noise ratios: the detection electronics for both methods pushes the state of the art, and signals are kept small by the danger

---

ment Foundation, the Sloan Foundation, and the IBM Corporation.

inherent in the use of more powerful lasers. Despite these difficulties Jarvis [20] has been able to generate a low resolution 64x64 pixel image in about 4 seconds.

## II. Principle of the Ratio Image Depth Sensor

The following discussion references Fig. 1, which shows a planar slice of a three dimensional system. As shown, an illuminating beam is projected onto a work area which is surveyed by a camera-like device. It is clear that the location of any point in the work space is uniquely determined by the intersection of a ray from the 'camera' and a ray from the 'projector'.

Suppose that the rays of projected light can be given some property P which varies monotonically across the beam, which is invariant under reflection, and which can be sensed by the special camera shown in Fig. 1. Then for each of many directions across its field of view the camera records the value of this property possessed by the projected beam where it is reflected to a camera pixel. Suppose for example that at the camera pixel corresponding to ray R the sensed value of the property P is V. If the reflecting surface were at a different location along the ray R the value of the property P sensed by the camera would be different from V. This allows the camera to generate an 'image' of the work scene which contains values identifying the 3D position of the reflecting surface observed at each pixel position.

But what optical property is to be used for the 'P' assumed in the preceding paragraph? The obvious simple properties such as intensity, color, or polarization of common light cannot be used, since all these can be changed considerably under reflection. However, except in very unusual cases, all factors (such as distance from and inclination to illuminating source, albedo of reflecting object, etc.) which determine the fraction of the incident light reflected to the camera are independent of the intensity of the incident light. This allows us to use the simple idea sketched in the preceding paragraph simply by taking the pixel-by-pixel ratio of two ordinary digitized images. Specifically, a first image is made with the scene illuminated by a beam of light which varies monotonically in intensity from one side to the other. (Such a beam can be formed using a slide projector and an appropriate graded neutral density filter.) Then a second image is made with a beam of uniform intensity. The two resulting intensity images are divided pixel-by-pixel. This division cancels out all factors (except the intensity of the incident light) which affect the intensity of the reflected light; the resulting quotient or ***ratio image***, contains (only) information about the location of surfaces within the 3D scene.

There are many considerations (such as choice of filter function, optimization of projector-camera geometry, etc.) which can be attacked theoretically. There is however one over-riding concern: can such a device really be made to work? The experiments described in the next section address this practical question.

## III. A Preliminary Depth Sensor Implementation

We have begun a series of experiments to measure many of the engineering parameters of the proposed Ratio Image Depth Sensor, e.g. the stability and definition which can be attained in the projected light beams and the relevant aspects of camera response, such as linearity, noise immunity, stability. The implementation described here allows us to make depth images quickly and with a minimum of computation, and thus to test our understanding of the physical and technical factors involved in the process. We note the results of some experiments on isolated components of the sensor, reproduce a depth image made with this implementation, and discuss the sources of experimental error in the measurements.

The key simplifying assumption in this implementation is that the change in depth is proportional to the change in the observed ratio along each ray from the camera through the work space. The calibration scheme to which this approximation leads is sketched in Fig. 2. A screen is placed perpendicular to the camera axis at a 'near' location ($z_N$) and a ratio image is made of this screen; the screen is then placed at a 'far' location ($z_F$) and a second ratio image is made of the screen. (A ratio image — the pixel-by-pixel quotient of two intensity images — behaves as any other image under ordinary image processing, and functions as a basic entity in any discussion of this technique.) These two vertical planes define the work area within the field of view of the camera. For any pixel $i,j$ it then follows by assumption that the measured depth $\bar{z}$ is

$$\bar{z}_{i\,j} = \frac{R_{i\,j} - R^N_{i\,j}}{R^F_{i\,j} - R^N_{i\,j}} (z_F - z_N) + z_N \tag{1}$$

where $R^N$ and $R^F$ are the ratio images of the near and far screens. Clearly systematic small distortions in the measured depth values must be anticipated; however, this implementation makes it possible to test the equipment under operating conditions and to estimate the practicality of more accurate (and more complicated) implementations.

The apparatus used in this implementation (shown in the block diagram of Fig. 3) consists of a slide projector, solid state television camera, and a VICOM image processor with a VAX 750 running Unix 4.2bsd acting as host. The VICOM has a firmware operating system which supports a large number of commands which operate on entire images in a television frame time. Images up to 512x512x12 bits deep are supported in all operations.

A software environment for image processing developed at the Courant Institute by Clark and Hummel [21] provides a UNIX shell that facilitates access to the VICOM. This shell is extremely flexible in its full implementation, making all normal shell facilities available to VICOM users and making the VICOM available to programs on the VAX. This shell is used in the present experiments primarily to pass files of commands to the VICOM for execution.

A Fairchild CCD-3000 camera equipped with a Fujinon-TV 25 mm f/1.4 lens provides a standard RS-170 video signal to the VICOM. A Matthey 4.25 MHz low pass video filter smoothes the output of the camera for sampling by the VICOM. The field of view is approximately 25° wide in the horizontal direction. The response of the camera at each pixel is proportional to the image intensity from zero to the maximum value of the output; however, with light levels corresponding to approximately six times the maximum output value regions of the image related vertically to an overloaded area are severely affected. The video signal is digitized by the VICOM in real time using an 8 bit A/D converter, but there is sufficient noise to allow averaging of successive images to acquire a 10 bit intensity image.

The Kodak Ektagraphic III B projector used in these experiments is equipped with an f/ 3.5 zoom lens (100 to 150 mm) and remote slide changing capability. Experiments have shown that the effect of defocussing of the 'ratio rays' through the work area does not represent a significant source of error, and that filters are repeatedly placed within 0.004 inch of the same location. Measurements of the temporal stability of the intensity of the unfiltered projected beam show a peak-to-peak variation of the intensity of 6% of the average brightness at a frequency of 120 Hz; this variation is apparently averaged by the camera and does not pose a problem. In addition there is a slow random variation (period of about 2 seconds typically) with a peak-to-peak amplitude of about 1% of the average brightness, and this could be significant in the present implementation. Dust and dirt on the filters represents a potentially large source of error.

The neutral density metal on glass filters [22] used in these experiments show a nominally linear variation in optical density along the length of the 1x2 inch (2.54x5.08 cm) filters. The isodensity contours form nicely straight lines across the filters, perpendicular to the direction of gradient change along the length of the filters. In the measurements reported here the ratio images were formed by dividing an image formed using a plain glass slide with an image formed with a filter which varied in transmissivity by a factor of 2 along its length. The ratio for this particular combination varies almost linearly across the 'ratio beam' projected on a flat screen nearly perpendicular to the beam.

The procedure for making ratio images is as follows:

1) digitize scene in ambient light
2) digitize scene lit by filter 1 (plain glass)
3) digitize scene lit by filter 2 (factor 2 variation in transmissivity)
4) form ratio $(I_1-I_A)/(I_2-I_A)$

In the actual measurements each intensity image is formed by averaging eight consecutive 8 bit images. The 512x512 images are spatially averaged by passing a hollow 3x3 convolution box over them twice, then they are subsampled to form 256x256 images (to conserve memory).

The depth images reported here were formed using a work space 30 cm deep centered 70 cm from the front of the camera. The lens of the slide projector was placed 142 cm to the left and 46 cm behind the front of the camera lens, and was directed towards the center of the work area. The depth sensor was calibrated by making ratio images of a matt white formica screen in the near and far locations;

the calibration process requires 3 to 5 minutes. All depth images were formed by making a ratio image of the scene (using the above procedure) and computing the measured depth $\bar{z}$ given by eq. 1. In the present implementation all computations are performed on the VICOM, and about a minute and a half is needed to form a depth image.

The scene shown in Fig. 4 is contrived to illustrate the differences between a depth image and a more usual intensity image. A sheet of heavy card stock stands toward the back of the work space defined above. Near the middle of the work space letters cut from construction paper stand on thin wooden sticks, which are hidden for the most part by a flat sheet of dark paper which has the letters IMAGE in paper of strongly contrasting reflectivity pasted on it.

An intensity image of this scene, illuminated by the slide projector and viewed by the television camera, is shown in Fig. 5a. The image was formed in the normal full ambient light of the laboratory. The shadows of the free standing letters DEPTH which appear on the back screen suggest the displacement of the letters from the background. The intensity of the letters DEPTH varies primarily because the letters are rotated about their vertical axes, and thus make different angles to the incident light; this strongly affects the intensity of the reflected light. In addition, the intensity of the light varies across the scene because the projector beam is brighter in the center than at the edges, the intensity falls off with distance from the projector, and the filter reduces the intensity toward the right side of the scene.

In the depth image of the same scene, Fig. 5b, brightness corresponds to closeness to the front plane of the work area, i.e. darker areas are more distant, with the exception that black areas indicate regions where depth information is not available due to shadows in the original image. (The bright white patches bordering the black are artefacts of the calculations and display look-up table.) The intensity information of the original scene is not present in the depth image; instead intensity is used to record distance in the depth image.

We have analyzed this depth image and similar ones for random variation from image to image. Typically consecutive images show an average deviation between corresponding pixels of between 1% and 4% of the depth of the work area, with the larger uncertainty being observed in dark regions of the intensity image; this is consistent with the analysis of the uncertainties given below. For this work area of 30 cm depth, 1% corresponds to 3 mm, 4% to 12 mm. The source of most of this noise is the random pixel-by-pixel variation in the digitized intensity images, although some variation could arise from changes in the ambient and projected light. Further work is required to identify the source of the uncertainties fully. In addition to the random noise, there is error in the depth measurements due to systematic error introduced by the simple method of calculating the depth.

## IV. Analysis of Experimental Uncertainty

The ratio image method requires that the digitized image accurately preserve the light intensities observed at the camera. However, any measurement is subject to uncertainty; in the present implementation sources of such error are fluctuation in the intensity of the projected beam, random noise in the camera/digitizing electronics, and variation in the ambient light. Additional effects which could degrade the performance of the depth sensor are non-linearities in the camera response, loss of resolution during calculations, and errors of approximation and inaccuracy in the calibration procedure. In this section we consider the effect of digitization noise on the resolution of the depth sensor.

A relation between depth resolution, ratio resolution and intensity resolution can be established by considering the case in which the ratio varies linearly across the beam. The intensity resolution is modeled using the function which has been measured for the noise under a number of typical operating conditions, namely $\Delta I = 0.5+0.004I$, where $\Delta I$ is the average deviation of the intensity and $I$ varies from 0 to 255. For example, for $I$ = 10, 20, 50, 100, and 200 units this results in a relative uncertainty $\Delta I/I$ of 5.4, 2.7, 0.9, and 0.65 %. This suggests that measurements should be made with the intensity as high as possible.

The ratio image is formed by dividing two intensity images, say $R = I_1/I_2$. The uncertainty in the resulting ratio is, in the case of small deviations which we are considering here, related to the uncertainty in the intensity images by

$$\frac{\Delta R}{R} = \frac{\Delta I_1}{I_1}+\frac{\Delta I_2}{I_2}. \tag{2}$$

For example, if $I_1$ varies from 100 to 200 linearly and $I_2$ is a constant 200, then $\Delta R/R$ varies from 1.55% to 1.3%.

The uncertainty in the ratio directly affects the uncertainty in the measured depth, and the size of this effect can be estimated by supposing that a typical ray from the camera, in traversing the workspace, will encounter variation in the ratio of approximately one half the total ratio range. For example, if the total range of variation of the ratio across the workspace is 0.50 (from 0.5 to 1.0 say), the ratio might range from 0.50 to 0.75, from 0.65 to 0.90, etc. along various rays from the camera through the work space. Taking 1.4% as a typical value for $\Delta R/R$ and 0.75 as a typical value of $R$, $\Delta R = 0.01$, or about 4% of the variation of $R$ along a ray through the workspace. The relation between ratio and depth can be taken to be linear for purposes of estimating uncertainties. Thus if the workspace is $D$ cm deep (corresponding to the 0.25 change in the ratio $R$), the average deviation in a depth measurement would be $0.04D$. For $D = 30$ cm this would be $\Delta D = 12$ mm.

The estimates given above refer to uncertainty for a single digitization. Since the noise is random, averaging of successive measurements will reduce the uncertainty in the average measurement. The averaging of neighboring pixels also can reduce random error, although this results in loss of definition at discontinuities. As an example, averaging four successive intensity images and one immediate neighborhood (9 pixels) would reduce the above $\Delta D$ from 12 mm to 2 mm.

The output of the camera varies linearly with light intensity within an experimental uncertainty of 2% in preliminary experiments. Since small variations from pixel to pixel in the proportionality constant divide out in making the ratio image, and small deviations in the zero offsets are cancelled when the ambient light is subtracted, the remaining potential source of systematic camera error lies in the possible existence of non-linear response. While the effect of these errors could be of magnitude comparable to the random noise, we have not yet observed any error traceable to non-linearities of the camera response.

The theoretical analysis of the error in depth images given above is based on favorable assumptions regarding the intensity of light reflected from the work scene to the camera, which depends on the intensity of the incident light (filter transmissivity, beam distribution, distance from projector) and on the surface (reflectivity, inclination to the light). We observe, without going into details, that the filter transmissivity varies by a factor of 2 (or more), the unfiltered intensity changes by a factor of 2, the reflectivity (ignoring specular reflections!) typically by a factor of 5 or 10, and the inclination of surfaces to the incident light contributes a factor of 2 reduction (corresponding to a surface at 60°) and very large factors when grazing angles are encountered. There are of course even more extreme instances (the eyes of a black cat in a coal bin) but usual work scenes should be within these bounds. The numbers given here suggest that intensities encompass a factor of about 80; fortunately unfavorable combinations seem to be rare, and it appears reasonable to construct a depth sensor with a smaller dynamic range. The dynamic range of the present implementation, about a factor of 10, is too small, and results in large error in poorly illuminated regions. This limited dynamic range represents the most serious engineering deficiency encountered in this preliminary implementation, and our work in the near future will aim to correct this.

References:

1. R.A. Jarvis, A perspective on range finding techniques for computer vision, IEEE PAMI-5, pp. 122-139 (1983).

2. A.C. Kak, Depth Perception for Robots, in Handbook of Industrial Robotics, S. Nof, ed; John-Wiley (to be published).

3. J.D. Boissonnat and F. Germain, A new approach to the problem of Acquiring Randomly oriented workpieces out of a bin, Proc. 7th IJCAI, pp. 796-802 (1981).

4. O.D. Faugeras, F. Germain, G. Kryze, J.D. Boissonnat, M. Herbert, J. Ponce, E. Pauchon and N. Ayache, Towards a Flexible Vision System, in Robot

Vision, A. Pugh, ed; IFS (Publications) Ltd., U.K. and Springer-Verlag, Berlin, Heidelberg, New York, 1983.

5. Y. Sato, H. Kitagawa and H. Fujita, Shape measurement of curved objects using multiple slit-ray projections, IEEE Trans. PAMI-4, pp. 641-646 (1982).

6. M. Oshima and Y. Shirai, Object recognition using three-dimensional information, IEEE Trans. PAMI-5, pp. 353-361 (1983).

7. R.C. Boiles, P. Horaud, and M.J. Hannah, 3DPO: a three-dimensional part orientation system, Proc. 8th IJAIC, pp. 1116-1120 (1983).

8. The White Scanner, Technical Arts Corp., PO Box 15390, Seattle, Washington, 98115.

9. J.T. Schwartz, Structured light sensors for 3-D robot vision, NYU Courant Institute Robotics Rpt. no. 8, 1983.

10. D. Rosenberg, M.D. Levine and S.W. Zucker, Computing relative depth relationships from occlusion cues, Proc. 4th IJCPR, pp. 765-769 (1978).

11. R. Bajcsy and L. Lieberman, Texture gradient as a depth cue, Comput. Graphics Image Processing, 5, pp. 52-67 (1976).

12. K. Ikeuchi, Numerical shape from shading and occluding contours in a single view, MIT AI Memo 566 (1980).

13. H.P. Moravec, Robot Rover Visual Navigation, UMI Res. Press, Ann Arbor, 1981.

14. B.K.P. Horn and B.G. Schunk, Determining Optical Flow, MIT AI Memo 572 (1980).

15. K. Ikeuchi and B.K.P. Horn, An application of the photometric stereo method, Proc. 6th IJCAI, pp. 413-415 (1979).

16. R. Ray, J. Birk and R.B. Kelley, Error Analysis of Surface Normals Determined by Radiometry, IEEE Trans. PAMI-5, pp. 631-645 (1983).

17. Y. Shirai and M. Suwa, Recognition of polyhedrons with a range finder, Proc. 2nd IJCAI, pp. 80-87 (1971).

18. D. Nitzan, A.E. Brain and R.O. Duda, The measurement and use of registered reflectance and range data in scene analysis, Proc. IEEE 65 pp. 206-

220 (1977).

19. R.A. Lewis and A.R. Johnston, A scanning laser rangefinder for a robotic vehicle, Proc. 5th IJCAI, pp. 762-768 (1977).

20. R.A. Jarvis, A laser time-of-flight range scanner for robotic vision, IEEE Trans. PAMI-5, pp. 505-512 (1983).

21. D. Clark and R. Hummel, VSH user's guide: a software environment for image processing, NYU Courant Institute Robotics Rpt. no. 19, 1984.

22. Dyn-Optics, 26098 Getty Dr, Laguna Niguel, CA 92677

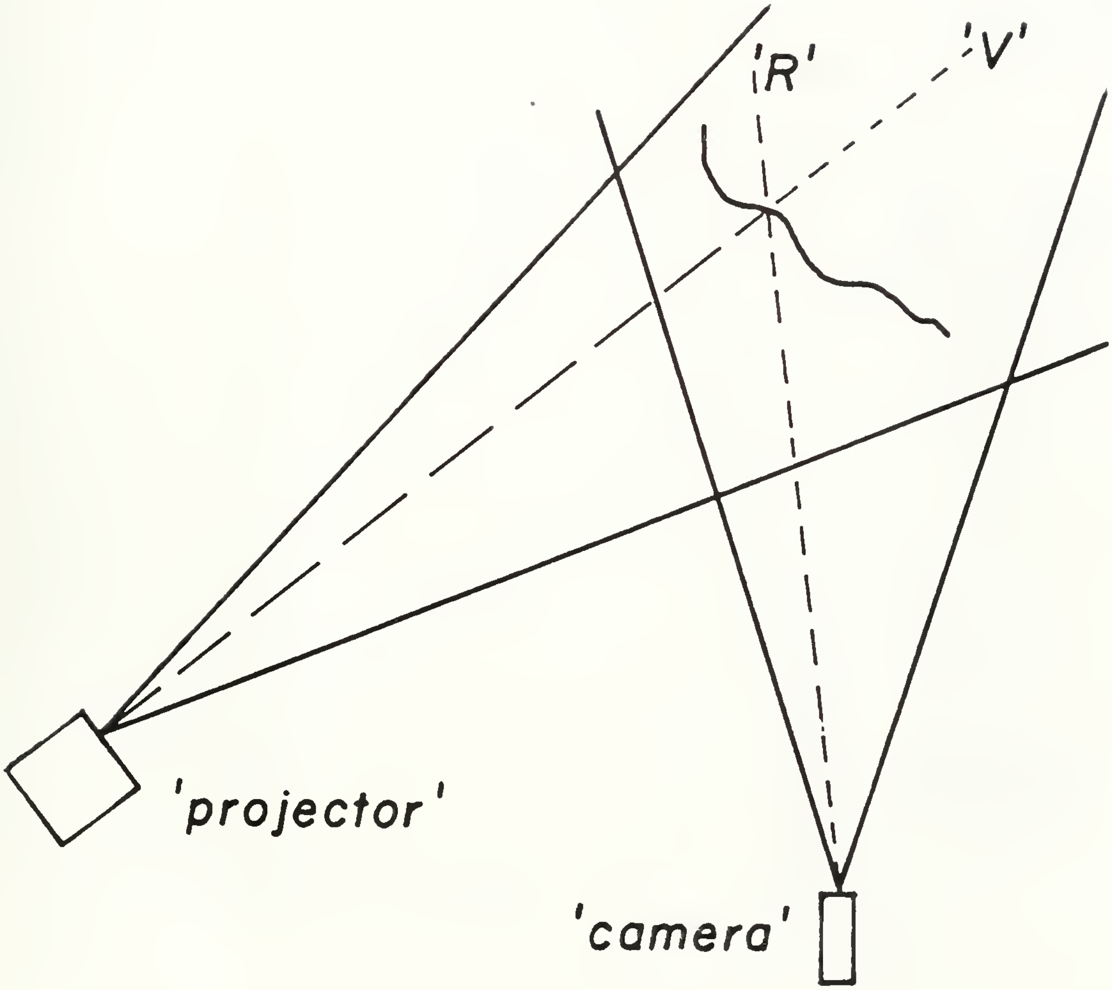


Fig. 1. Schematic representation of a general optical triangulation scheme. A point in the workspace is uniquely determined by the intersection of a ray from the 'projector' and a ray from the 'camera'.

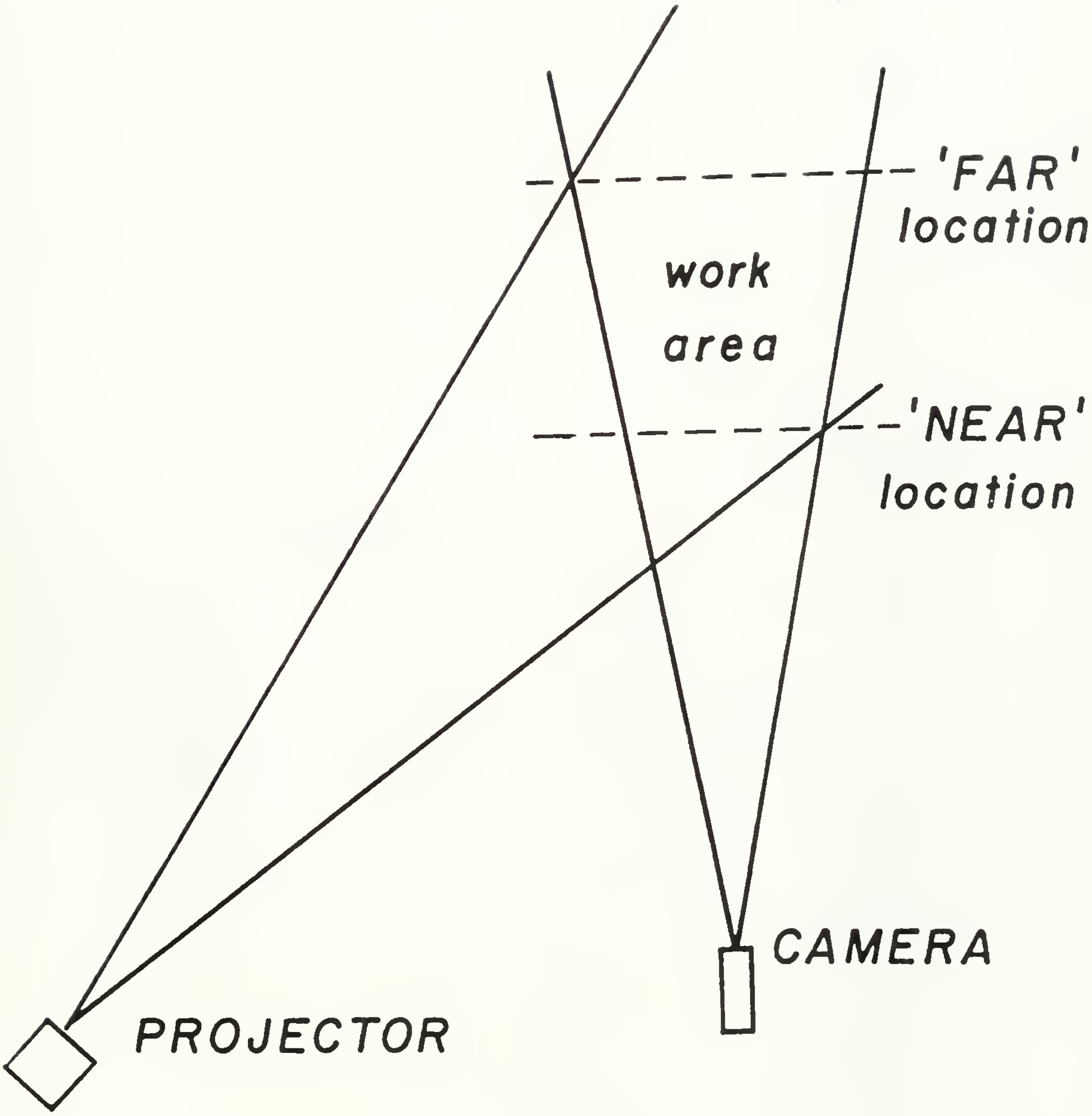


Fig. 2. Preliminary implementation of the ratio image depth sensor. Ratio images of a vertical screen placed at the 'near' and 'far' locations are formed successively, resulting in ratio values for the 'near' and 'far' locations at each pixel location. Depth is interpolated as a linear function between these end values for each camera ray through the workspace (see eq. 1).

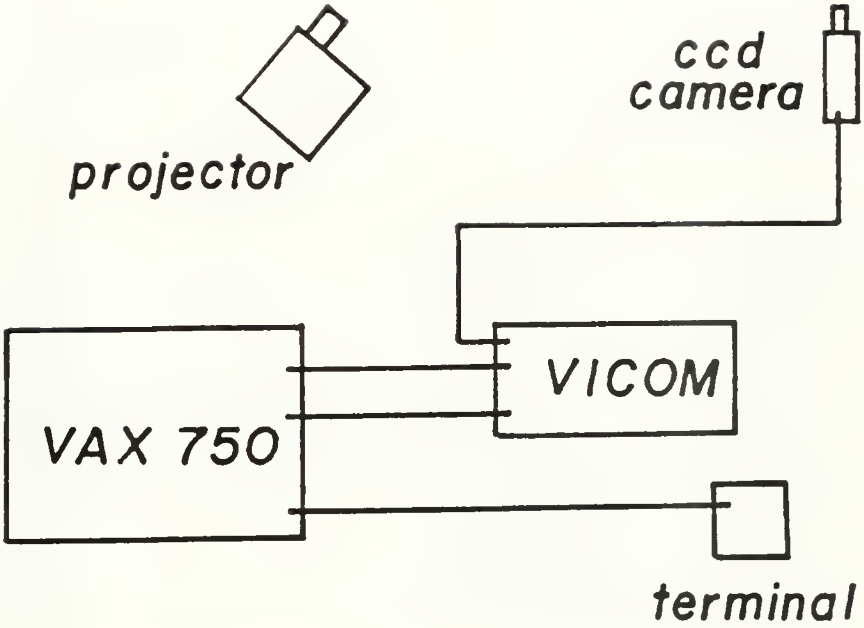


Fig. 3. Block diagram of the equipment used in these experiments. The VICOM and VAX are connected by both 9600 baud serial line and high speed parallel line for the dma transfer of images.

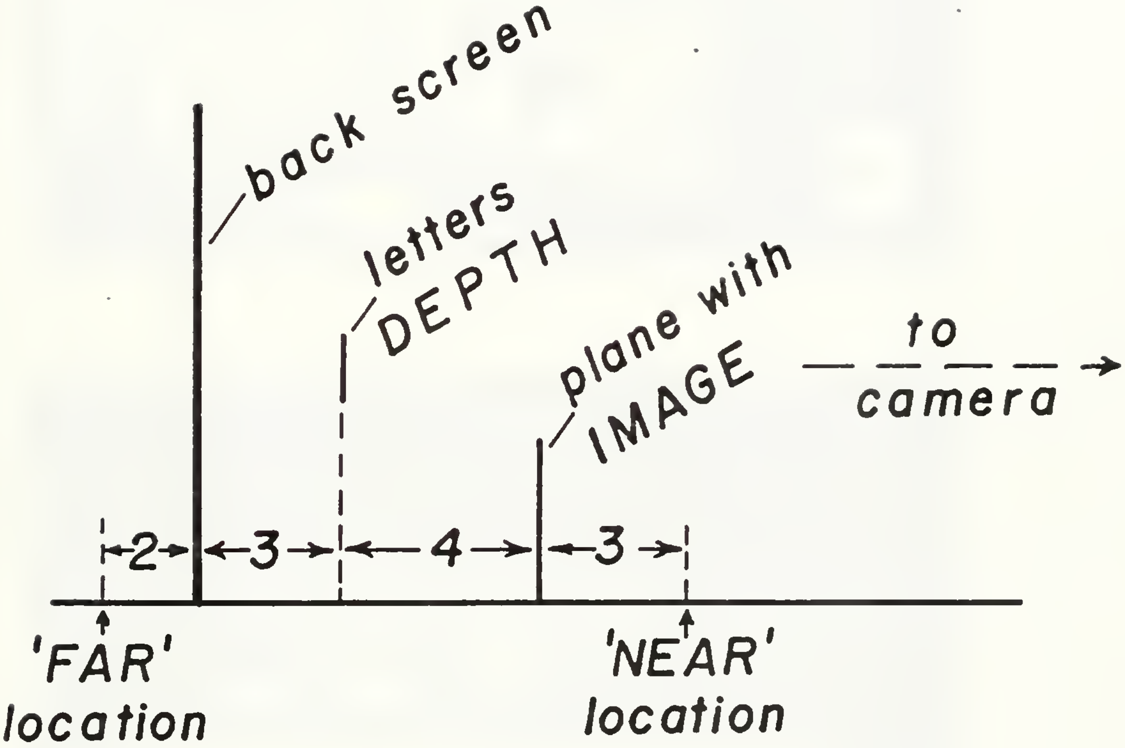


Fig. 4. Schematic representation of the demonstration scene used to make the images shown in Fig. 5. The letters DEPTH, cut from stiff paper, stand freely near the middle of the workspace, which is 30 cm deep.

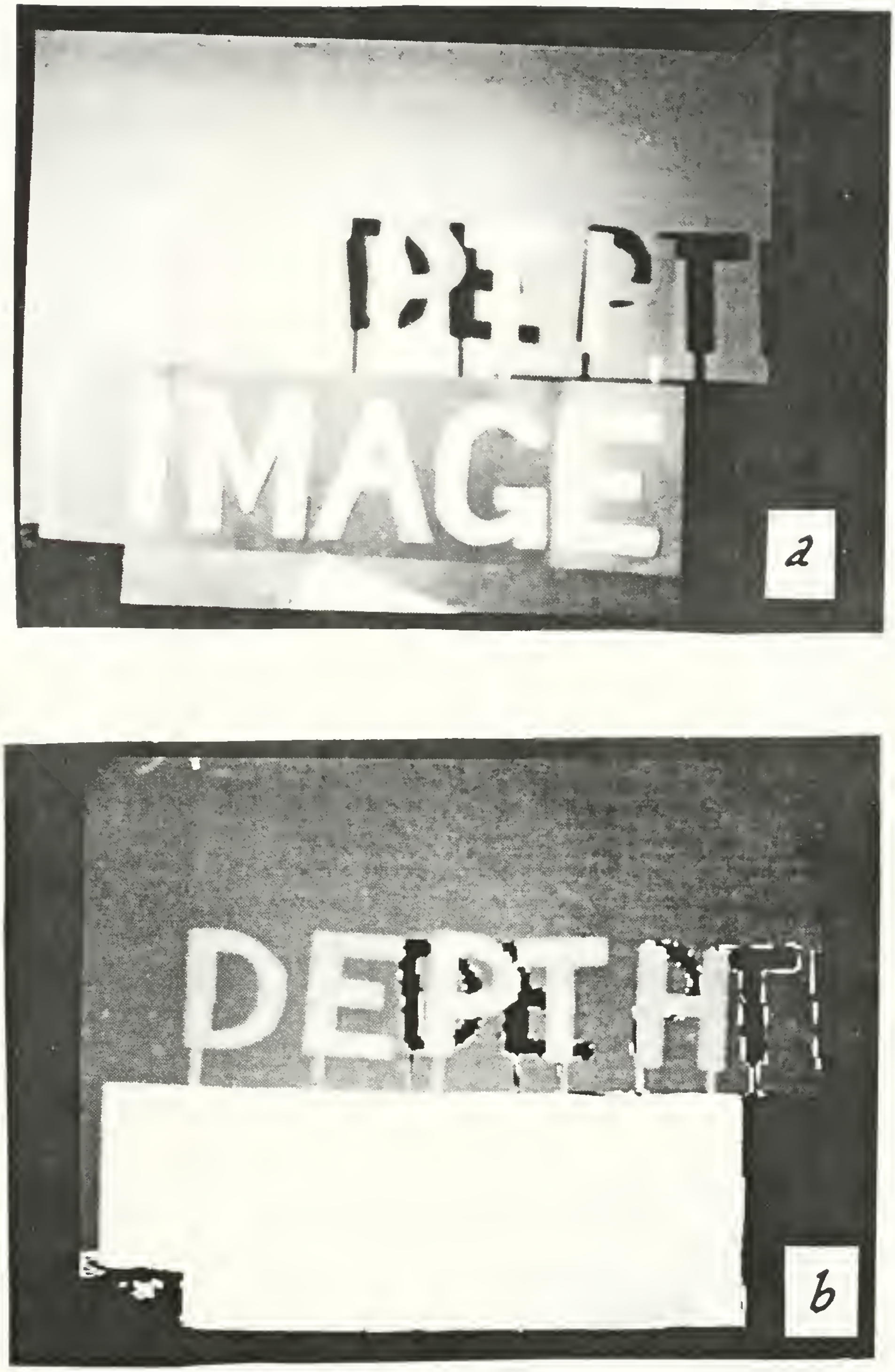


Fig. 5. Intensity (a) and Depth (b) images of the scene sketched in Fig. 4. The intensity image is a normal view of the scene illuminated by the slide projector. In the depth image (below) the intensity codes distance from the camera, brighter being closer, darker being more distant.

C1

NYU/CS/TR-124
Bastuscheck

Preliminary implementation

C1

NYU/CS/TR-124
Bastuscheck

Preliminary implementation
of a ratio image depth...

| DATE DUE | BORROWER'S NAME |
|---|---|
| ~~SEP 27 1985~~ | ~~Chin Woo Yoo~~ |
| | |
| | |
| | |